

Earth Science Data Search: Why So Difficult?!@#%?

- A FROST-y Solution

Use Case

- Try to find all the aerosol data for the May Chaiten eruption: gridded data, high-resolution swath, model output, experimental products...
 - Candidate Sources: ASDC, CNES, GES DISC, DataFed, MRDC, NCAR, P.I.s...

Directories Galore

- GCMD
- GEOSS
- GOS
- Mercury
- DataFed
- They're all different: holdings, access, structure

Inventories Galore

- ECHO
- Mercury
- DataFed
- Data Centers
- All have limited inventory-level holdings, and even more differences in access and structure than directories

What Do We Want?

- Access to datasets in multiple directories
- Access to files in multiple inventories
- A standards-based solution
- The ability to search from analysis and other clients (i.e., non-special-purpose search clients)

The FROST Solution

- **Federated Inventories** supply search results
- **Recursive Open-Search** dataset results point to open-search description docs for inventory level search
- **Tools** Off-the-shelf tools are provided to inventories to index and supply search results

Open Search

- www.opensearch.org
- a collection of simple formats for the sharing of search results
- OpenSearch description document describes a search engine so that it can be used by search clients (incl. Firefox and IE)
- OpenSearch response elements can be used to extend existing syndication formats, such as RSS and Atom, with the extra metadata needed to return search results.

Recursive OpenSearch

- OpenSearch search for datasets returns pointers to OpenSearch description documents for each dataset
- OpenSearch description document allows client to execute OpenSearch of that dataset's inventory
- XSL might even be used to generate a search form in the browser

Minimalist Scenario

- User searches for datasets based on keyword
- Returns a page with links to dataset OpenSearch description documents (OSDD)
- User clicks on one dataset's OSDD
- Browser retrieves XML and renders as search form
- User fills in space / time and searches for individual files

**No Special
Client Needed**

Machine-to-machine Scenario

- Application searches directory based on keyword, or...
- Includes links to dataset OSDDs in configuration file
- Application retrieves OSDDs, ignoring XSL reference
- Application executes search against individual inventories with user or application space / time constraints
- Application integrates XML inventory results

FROSTools

- A simple database engine
- A simple Web interface to parse and respond to OpenSearch requests
- The ability to index (by time and space) well-formed data files
 - KML, NetCDF/CF1, HDF-EOS, GRIB
 - ...or ancillary XML metadata files

What Do We Need to Make?

- XSL to render OSDDs as search forms
- End-to-end Proof of concept
 - Directory-level search: use GCMD?
 - Inventory search: gateway in front of ECHO
- FROSTools
- FROSTool packaging for easy install and configuration